Toxicology Excellence For Risk Assessment
1250 Ohio Pike, Suite #197
Cincinnati, Ohio 45102-1239

June 3, 2019

To: Science Advisory Board, U.S. Environmental Protection Agency

From: Michael L. Dourson, PhD., DABT, FATS, FSRA
Director of Science, Toxicology Excellence for Risk Assessment (TERA)
President, Toxicology Education Foundation

Re: Public Comments

Dear Dr. Honeycutt and Colleagues,

I offer the following comments on issues before your esteemed committee.

- Although it no doubt can be improved and abuse must be safeguarded, the **Science Transparency Rule** is necessary from a risk assessment perspective. Why? An example may be instructive. We currently have one chemical in judicial review based in large part on an observational epidemiology study suggesting IQ deficits in children. If these effects were true, this would be of grave concern. However, the study is in contrast to the wealth of other guideline-study data and it's suggested hypothesis has been tested in experimental animals and found not to be supported. Moreover, the published study is missing up to 35% of the data stated to be otherwise available as demonstrated in a recent SOT poster (attached) and the authors have not released their data, even redacted, to EPA, despite being asked twice, despite the study being publicly funded in part, and despite EPA's ability to handle otherwise confidential data routinely. Determining a risk assessment position with such data is highly uncertain and easily challenged by other scientists. Not using data in such situations is a reasonable policy decision.
- EPA has published progressive **Guidelines for Carcinogen (2005) and Noncancer (2002) Assessment** and is acclaimed by all to be a leader in these areas as demonstrated by extensive use and citation. However, risk assessment science has matured from the date of these last publications. Concepts such as Data-derived extrapolation factors (EPA, 2014), Bayesian Benchmark Dose (BMD), risk above the safe dose, high throughput toxicity screens, adverse outcome pathway (AOPs), and new thinking on mode of action (MOA) all can, and should, be integrated into revised guidelines. Several of these topics can easily and quickly be incorporated. Other topics will likely require additional study, debate, and judgment. See a recent poster at the NAS (2019) for one such consideration on dual MOA/AOP for the cancer findings of acrylamide (attached).
- **Per- and Polyfluoroalkyl Substances (PFAS) Action Plan** is important and as such should incorporate any newly available data. Recent findings of PFOA administration to 43 cancer patients with good renal clearance and outstanding kinetic follow-up appear to contradict observational studies of long half lives in humans and suggest different approaches to the highly disparate (~750-fold) safe dose assessment of different governments. EPA can take the lead here in both a revised safe dose assessment and in research on human PFOA/S kinetic clearance to

address a more appropriate animal to human extrapolation. A recent paper submitted on this topic can be accessed at: https://www.tera.org/about/news.html (see 5-14-19 story).

- **EPA's Proposed Waters of the U.S. Rule** is important and more or less consistent with the definition of *navigable waters* in the act. Although the desire to extend the rule to vernal waters, swamps, or marshes as some suggest is not unreasonable from several viewpoints, it would necessitate the application of ambient water quality criteria to such waters. The development of these criteria assumes that humans drink 2 liters of water per day and eat 6 to 20 grams of fish per day from such waters. From a practical occurrence and risk assessment perspective this makes little sense. The SAB might consider this risk perspective in their discussions as appropriate.

If time allows, I would be happy to answer any questions.

Sincerely,

Michael L. Dourson, Ph.D., DABT, FATS, FSRA
Director of Science

TERA

A Commentary on Some Epidemiology Data for Chlorpyrifos

Michael Dourson, PhD, DABT, FATS, FSRA
Bernard Gadagbui, MS, PhD, DABT, ERT
Chijioke Onyema, MPH
Patricia McGinnis, PhD, DABT

Toxicology Excellence for Risk Assessment (TERA)

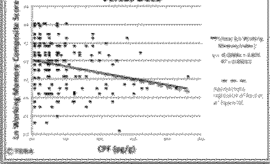
1

Results

- Figure 1 is a replication the Rauh et al. (2011) Figure 1A. This replication seems reasonable from a comparison of where the linear regression line lies in relationship to the approximate regression of Rauh et al.
- ~35% of the data stated in the publication in Rauh et al. (2011, page 1197) are missing in Fig. 1.
- Also not included are high dose data (e.g., see reference to 63 pg/g on Rauh et al. page 1198, column 2, which is not found on Figure 1). Apparently also missing is one child with a value of 32 pg/g (see stated CPF range in Rauh et al., Table 1).

2

Figure 1. Ln Working Memory Index Versus Dose



3

Results Summary

- The bottom line is that evidence of effect for Full-Scale IQ does not exist when data are presented as in Figure 4.
- Working Memory shows evidence of a negative association (Figure 2), but this evidence is problematic due to missing 35% of the data, including the highest exposed individual that an author states "was a highly influential observation (outlier) and drastically impacts inference." (Whyatt, 2015)
- Data from 3 other high dose individuals are also missing.
- Overall, the lack of raw data from this study makes confirmation of the authors' results impossible.

4

Discussion

- The most significant challenge, by far, in any reanalysis of the Rauh et al. (2011) study is the absence of data to conduct a credible replication. For example, Rauh et al. (2011) state that:
  - "Of 725 consenting women, 535 were active participants in the ongoing cohort study at the time of this report, and 265 of their children had reached the age of 7 years with complete data on the following: a) prenatal maternal interview data, b) biomarkers of prenatal CPF exposure level from maternal and/or cord blood samples at delivery, c) postnatal covariates, and d) neurodevelopmental outcomes."
- However, Rauh et al. (2011) show a series of 5 graphs, two of which (Figures 1A and 1E) show ~35% or ~15% missing data points, respectively. Neither of these graphs include the data points from highest stated cord blood exposures of 32 or 63 pg/g.

5

Discussion Continued

- Our analysis from published graphs is a rudimentary way to obtain the data of Rauh et al. (2011), because data points may often overlay one another in published figures. Still, the Rauh et al. (2011) do not plot chlorpyrifos exposures greater than 25 pg/g.
- Not surprisingly, as study co-sponsors, EPA asked Rauh et al. (2011)for these data–twice. Such requests seem reasonable since:
  - "This study was supported by the National Institute of Environmental Health Sciences (grants ...), the U.S. Environmental Protection Agency (grants ...), the Educational Foundation of America, the John and Wendy Neu Family Foundation, the New York Community Trust, and the Trustees of the Blanchette Hooker Rockefeller Fund."

6

Conclusion

- Rauh et al. (2011), has been cited as showing a statistical association between CPF exposure and intelligence.
- An analysis of the published figures shows that up to 35% of the data appear to be missing, including high dose data, inclusion of which would appear to mollify or obviate the effect as stated by the authors.
- The associations are lessened or no longer apparent when different logarithmic assumptions were used in the reanalysis.
- The data, generated in part by public funds, have not been made available for independent review.
- These shortcomings make it difficult to use this study in any risk assessment for chlorpyrifos.

7

Abstract

- Rauh et al. (2011) claimed statistically significant associations for neurological effects in children at age seven after potential very low exposure to chlorpyrifos (CPF) during pregnancy detected in a single cord blood at birth, with 42% non-detects.
- Although the underlying data have not been made available, we were able to extract data of Figures 1A and 1E in Rauh et al. (2011). Our analyses showed ~35% of the data in Figure 1A and ~15% of data in Figure 1E were missing from 265 children.
- Our analysis of extracted data does not suggest any evidence of an effect on Full-Scale IQ (Figure 1E), and less of a negative association (reduction) in Working Memory Index (Figure 1A).

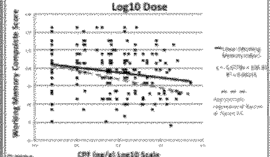
8

Results Continued

- Figure 3 attempts to replicate the findings of Rauh et al. (2011), specifically their Figure 1E (Full Scale IQ).
- This replication is not as close as in Figure 1. Again, compare where the regression lines lie in relationship to high dose points.
- As in comparison of Figures 1 and 2, some of the data stated to be available in Rauh et al. (2011) are missing (in this case approximately 15%).

9

Figure 2. Working Memory Index Versus Log10 Dose



10

Blank

11

Discussion Continued

- Although negative neurological associations are reported in the Rauh et al. (2011) when a plot of Working Memory Composite Scores are normalized by their natural logarithm and plotted with dose, this manner of data display is not the only one possible.
- Another, and more standard approach, would be to plot the unadjusted scores, which are already normally distributed in the human population, against the logarithm of dose.
- When the results of Rauh et al. (2011) are plotted in this way, a reduced association is found. For example, a comparison of Figures 1 and 2 will show that the negative trend of Figure 1 for the Working Memory is less in Figure 2.

12

Discussion Continued

- A number of other items about this study are noted:
  - Rauh et al. (2011) Figure 1 E has more data points than Rauh et al. (2011) Figure 1A, but the former depends on the latter.
  - Rauh et al. (2011) did not collect cord blood samples in 12% of the population. Blood made collected for only 34% of mothers and children.
  - The hypothesis from Rauh et al. (2011), that neurological effects occur at doses lower than cholinesterase inhibition, has been tested in experimental animals and is not supported.
  - The metabolite causing toxicity irreversibly binds to cholinesterase in the blood. So bound, it is not expected to affect fetal neurological development at levels lower than those not causing blood inhibition.
  - Systemic availability or brain cholinesterase inhibition has not been seen at levels comparable to safe doses (Marty et al., 2012).
  - The specific results described by Rauh et al. (2011) are not reproduced in other epidemiology studies (Burns, 2016).
  - Scientists with USDA, EPA's SAP, and OMB also have substantive questions about the Rauh et al. (2011) work.
  - The IQs in the control and exposed groups of Rauh et al. (2011) both approximated 98, and were not statistically significantly different.

13

Blank

14

Background

- Organophosphates are beneficial against Zika virus and malaria.
- Scores of studies suggest that sentinel effect of chlorpyrifos is cholinesterase inhibition at similar dose in animals and humans.
- Rauh et al. (2011) claimed neurological effects in children associated with much lower prenatal levels of chlorpyrifos, suggesting an alternative hypothesis, that the sentinel effect for chlorpyrifos should be based on neurological effects rather than cholinesterase inhibition.
- The purpose of this analysis is to consider this hypothesis by reviewing the underlying epidemiological basis.

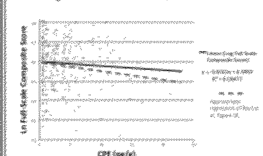
15

Results Continued

- Figure 4 reflects the data from Figure 3 plotted with the response unadjusted and the dose in logarithmic units (log_{10}), and shows no effect on Full Scale Composite score when compared with Figure 3.
- As with Figure 2, the y-axis is not compressed in Figure 4, eliminating the subtle visual effect of downward trend due to this compression in the y-axis of Figures 1 & 3.
- The R^2 is for Full Scale IQ is even smaller than for Working Memory, suggesting that chlorpyrifos is not a predictor of the outcome.

16

Figure 3. Ln Full Scale IQ Versus Dose



17

Blank

18

Discussion Continued

- When a similar analysis is performed for Full Scale IQ (or Composite Score) the slight negative trend of Figure 3, which is a representation of Rauh et al. (2011) Figure 1E, disappears; compare Figures 3 and 4.
- What about including the missing high dose data?
- Adding the two high dose data points described in Rauh et al. (2011) to figures 2 and 4 and supposing only average responses further decrease the negative slopes, but only slightly (data not shown). This indicates even less of an effect, if any, from chlorpyrifos exposure.

19

References & Notes

- Rauh et al. (2011). Seven-Year Neurodevelopmental Scores and Prenatal Exposure to Chlorpyrifos, a Common Agricultural Pesticide. Environmental Health Perspectives. Volume 119: 1196-1201.
- Memo to: Carol Christensen, Ph.D. From: Robin M. Whyatt, DrPH, Date: April 9, 2015.
- USEPA. 2014. Chlorpyrifos: Revised Human Health Risk Assessment for Registration Review.
- Marty et al. Cholinesterase inhibition and toxicokinetics in immature and adult rats after acute or repeated exposures to chlorpyrifos or chlorpyrifos-oxon.
- Burns, C.J. 2016. Reproducibility is critical for determining scientific validity.

20

Blank

21

Methods

- Figures 1A and 1E of Rauh et al. (2011) were extracted and the results entered into Excel. Consistent Rauh et al. (2011, page 1198), ~80% of zero or non-detectable values were assigned a chlorpyrifos level of 0.5 pg/g; ~20% of them assigned 1.0 pg/g.
- The results were then plotted as natural logarithm-adjusted response versus dose (as per Rauh et al., 2011), and as un-adjusted response versus log_{10} dose using Excel software.
- 4 high dose data points were missing in both graphs; authors' response to EPA states the highest dose "was a highly influential observation (outlier) and drastically impacts inference." One subject had no data; data of two others were not plotted.

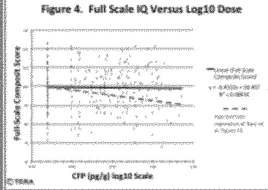
22

Results Continued

- Figure 2 reflects the data from Figure 1 plotted with the response unadjusted and the dose in logarithmic units (log_{10}).
- Figure 2 shows a reduced effect on Working Memory Index when compared with Figure 1, found by comparing where the regression lines lie in relationship to high dose data points, indicating that the way data are presented may have an effect on interpretation.
- The R^2 for regression lines in both Figures are very small, which indicates that chlorpyrifos does not explain the data variability (i.e. the scatter in these data).

23

Figure 4. Full Scale IQ Versus Log10 Dose



24

Blank

25

Discussion Continued

- Whether one method of plotting these data is superior to another may be important, but a strong association should not be affected.
- A more appropriate approach to confirm our findings would be to access the underlying data. This would enable us to adjust for confounding factors and discuss our results in more statistical terms, thus allowing a stronger statement on any statistical significant association.
- Underlying data would also allow a refined analysis from a simple linear approach to a manner similar to that shown by Rauh et al. (2011) as a smooth cubic spline curve.

26

Blank

27

Funding

- The origin of this work came about from briefings of MLD as a U.S. Environmental Protection Agency senior advisor in 2017.
- Afterwards, DowAgro Sciences funded, in part, an independently developed report by Toxicology Excellence for Risk Assessment (TERA), which confirmed the findings of EPA staff.
- The development and presentation of this poster has been supported by the Internal Development Reserve funds of TERA.

28

ED_006319_00004789-00003

Evidence-Based Dose Response Assessment for Thyroid Tumorigenesis; A tale of two MOAs... ...and what can be done about it

Michael Dourson[1], Rick Hertzberg[2], Bruce Allen[3], Lynne Haber[1,4], Ann Parker[1,5], Oliver Kroner[1,4], Andy Maier[1,4], and Melissa Kohrman[1,5]

[1]Toxicology Excellence for Risk Assessment, [2]Emory University, [3]Independent Consultant, [4]now at University of Cincinnati, [5]now at City of Cincinnati, [6]now at Cardno ChemRisk

1

Hazard Identification 1

- Acrylamide not shown to cause cancer in humans (Marsh et al., 2007; Swaen et al., 2007; Mucci & Adami, 2005).
- Six long-term experiments in rats show tumors (Johnson et al., 1986; Friedman et al.,1995; and NTP, 2012). All experiments used the same strain of rats.
- Two long term studies also available in mice (NTP, 2012).
- Four long term studies are also available in mice and rats for glycidamide, a known metabolite of acrylamide in vivo.

TERA

2

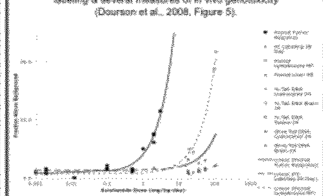
Mode of Action 1

- Acrylamide is genotoxic, but not directly mutagenic. Acrylamide also causes growth stimulation and oxidative stress, the latter of which can lead to mutations and other genotoxicity.
- Mutagenicity and genotoxicity from acrylamide exposure are not dose response concordant—only seen at doses higher than those that caused tumors (Figure 1).
- Moreover, in vivo mutations do not occur in the rat mammary gland or testes (Table 1 and Figure 2 of Mei et al., 2010).
- Accordingly, it is unlikely that the tumors evoked are solely caused by either mutagenicity or genotoxicity. However, unmeasured mutagenicity might be occurring at low doses and might be responsible for some of the low dose tumors (Figure 2).

TERA

3

Figure 2. Multiples above background for thyroid tumors, DNA labeling & several measures of in vivo genotoxicity (Dourson et al., 2008, Figure 5).



TERA

4

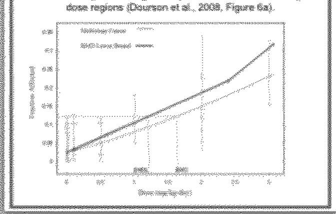
Dose Response Assessment 1

- EPA (2005) suggests "decoupling" data when several modes of action occur in different parts of the dose response curve. Thus,
  - a mutagenic, non-threshold, linear mode of action may be occurring at doses of less than 1 mg/kg-day, and
  - a growth stimulation, threshold, non-linear mode of action likely dominates at doses in excess of 1 mg/kg-day.
- EPA (2005) suggests that selection of a point of departure be close to the lower range of data of interest. Since controls had dose, a BMR lower than 5% is reasonable; a 2% extra risk is:
  - ~double the background rate of the pooled data, and
  - 1/3 of the highest low dose response rate of ~6%.
- Thus, a 2% BMR is comfortably within the interpolation range allowing for a stable estimate for the point of departure.

TERA

5

Figure 3a. Multistage model fitted to pooled-all thyroid tumor data, showing little change in slope between the low and high dose regions (Dourson et al., 2008, Figure 6a).



6

Abstract

- Acrylamide is commonly found in various foods.
- Cancer studies in rats report increases in many types of tumors. We find that thyroid tumors are most sensitive because they are repeatable, relevant and potent.
- Both mutagenicity & growth-stimulation occur; mutagenicity determines low-dose; growth stimulation at higher doses.
- We "decouple" these MOAs and determine that the probit model best reflects the underlying "decoupled" biology.
- We identify a cancer slope factor of 0.030 (mg/kg-day)$^{-1}$ for low dose, and a RfD of 0.05 to 0.02 mg/kg-day for high dose.

TERA

7

Hazard Identification 2

"Risk analysis shall be based on the most sensitive study deemed to be of sufficient quality" (California, 2011).

A risk assessment perspective of "most sensitive study" for determining the risk in humans includes studies that are:

- Repeatable. A sensitive study shows effects that are found in comparable studies of the same or other species/strains.
- Relevant. A sensitive study shows effects in one species that are expected to be found in humans.
- Potent. A sensitive study shows effects that occur at the lowest dose or concentration.

Most sensitive study ≠ only the most potent study

(See Appendix to this presentation)

TERA

8

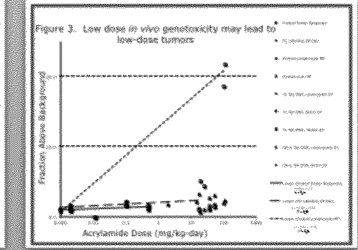
Mode of Action 2

- The weight of scientific evidence supports growth stimulation as contributing to thyroid tumors.
  - Khan et al. (1999) showed statistically significant morphological changes in the thyroid consistent with stimulation after acrylamide for 2 or 7 days.
  - Lafferty et al. (2004) observed three measures of growth stimulation in the thyroid after acrylamide for 7, 14 & 28 days; DNA labeling was statistically significant by pairwise comparison. Chico-Galdo et al. (2007) results suggest that this DNA labeling was growth stimulation.
  - Friedman et al. (1999) showed thyroid hormones to be statistically significantly decreased in males after 28 days of exposure (trend test & pairwise comparison). Females appeared to be affected, but less so.
  - Johnson et al. (1986) showed thyroid hyperplasia in both male and female rats after 2 years of exposure (statistically significantly-trend test).
- Comparison with Table 3 of EPA (1998) also supports this choice.

TERA

9

Figure 3. Low dose in vivo genotoxicity may lead to low-dose tumors



TERA

10

Dose Response Assessment 2

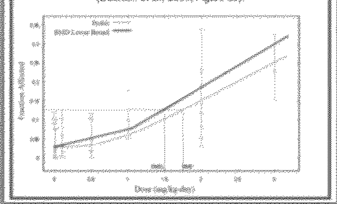
Using EPA (2005) guidelines, we compared different mathematical models to fit these "decoupled" data.

- Multistage model did not fit "decoupled" data as well (Figure 3a).
- Weibull model fit these "decoupled" data well with fixed power of 2, but EPA software did not allow a positive value for control doses with power unfixed; controls animals had a positive dose (Tweddle et al., 2004).
- Probit model fit these "decoupled" data well, showing a linear response for tumors in the low dose range and a curvilinear upward trend for tumors from growth stimulation in the high dose range (Table 4; Figure 3b).
- When low dose "decoupled" data only were considered, a weighted linear regression and a multistage model both confirmed the use of the probit model to generate a point of departure for the low dose extrapolation (Table 5; Figure 3c).

TERA

11

Figure 3b. Probit model fitted to pooled-all thyroid tumor data, showing differing slopes between the low and high dose regions (Dourson et al., 2008, Figure 6b).



12

Background

- Previous assessments were found to be generally conservative, "screening level" assessments.
- We undertook a more comprehensive risk assessment that benefited from the following:
  - new science on acrylamide's toxicity and its mode of action; and
  - new guidelines for the evaluation of cancer risk developed by the U.S. Environmental Protection Agency (EPA, 2005).
- Key determinants for evaluating tumor response are:
  - consistently observed tumors that can be statistically combined among studies using EPA research and guidelines, and
  - probable relevance to humans from experimental animals (note: when little is known such relevance is assumed).

TERA

13

Hazard Identification 3

- Thyroid tumors in rats exposed to acrylamide were repeatable and statistically significant in all six experiments (Table 1).
- Although, scientists have not identified any chemical that has caused thyroid tumors in humans and the rat thyroid is different from the human thyroid, we conclude that these rat thyroid tumors could not be discounted. Therefore, they are considered as relevant to humans.
- The type of thyroid tumors formed in rats is generally recognized as resulting from growth stimulation and/or mutation, and these modes of action also operate in humans.

TERA

14

Mode of Action 3

- Tumors evoked by acrylamide exposure were generally benign, occurred late in life, and were more often in hormonally-active organs, in all six rat experiments.
- Such tumor appearance is more consistent with manners of tumor formation that are not from direct mutation. These observations also mean it is unlikely that direct mutations are causing all of the tumors in these experiments.
- Thus, both a mutagenic and non-mutagenic manners of tumor formation are likely to contribute to thyroid tumors (Table 2).

TERA

15

Table 2. Key events, related evidence and plausibility of tumor modes of action.

[illegible]	[illegible]	[illegible]	[illegible]

TERA

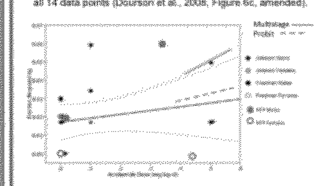
16

Dose Response Assessment 3

- The probit model Benchmark dose (BMD) of 2% extra risk is 0.81 mg/kg-day, associated with a slope of 0.025 (mg/kg-day)$^{-1}$.
- Slope value adjusted by 1.2 for known kinetic rat and human differences; not further adjusted for dynamic variability because:
  - Williams (1995) states that thyroid tumors in humans do not form in the presence of mutagens if TSH-stimulated growth is prevented;
  - EPA (1998) considers an adjustment factor of 1 for chemicals having a growth stimulation mode of action, unless specific data suggest otherwise;
  - Analysis of published studies comparing human & animal tumor slope factors show that the most likely value for an overall factor is roughly 1-fold;
  - Goodman and Wilson (1991) consider the best estimate of the interspecies factor to be log normally distributed around a value of 1.
- The adjusted slope factor is 0.030 (mg/kg-day)$^{-1}$.

17

Figure 3c. Weighted linear regression on low-dose, pooled data with 95% confidence curves for the model. Data jittered to show all 14 data points (Dourson et al., 2008, Figure 6c, amended).



TERA

18

TERA

Mission is to support the protection of public health by:

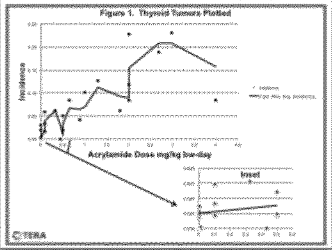
- Developing, reviewing and communicating risk assessment values and analyses;
- Improving risk methods through research; and
- Educating risk assessors, managers, and the public on risk assessment issues
- TERA is a 501c3 nonprofit organization

19

Table 1. Thyroid tumors in Johnson et al. (1986), Friedman et al. (1995), & NTP (2012) drinking water studies.

[illegible]	[illegible]	[illegible]	[illegible]

20



TERA

21

TERA

22

22

Table 4. Probit model estimates of slope factors (SF) for rat thyroid tumor data. BMD/L values in mg/kg-day; SF values in (mg/kg-day)$^{-1}$ (Dourson et al., 2008, Table 6).

Data set	BMD 0.02	SF at BMD 02	BMDL 0.02	SF at BMDL 0.02
Pooled female	0.82	0.024	0.67	0.030
Pooled male	0.97	0.021	0.58	0.034
Pooled all	0.81	0.025	0.69	0.029

	BMD 1.0	SF at BMD 10	BMDL 1.0	SF at BMDL 1.0
Pooled female	1.8	0.057	1.5	0.069
Pooled male	1.7	0.059	1.3	0.079
Pooled all	1.7	0.057	1.5	0.067

TERA

23

Table 5. Comparison of 2% BMD and BMDL from various models for acrylamide induced thyroid tumors in rats & projected slope factors.

Data Set	Model	BMD_{02}	SF BMD_{02}	$BMDL_{02}$	SF $BMDL_{02}$	BMD/BMDL
All pooled tumors	Multistage	0.39	0.052	0.23	0.087	1.7
Low-dose pooled tumors	Multistage	0.89	0.025	0.27	0.068	3.3
All pooled tumors	**Probit**	**0.81**	**0.025**	**0.69**	**0.029**	**1.2**
All pooled tumors	Weibull	0.82	0.024	0.72	0.028	1.1
Low-dose pooled tumors	Weighted linear regression	0.92	0.022	0.33	0.061	2.8

P-values for all regressions exceed 0.1 and thus are acceptable. AICs and residuals for the multistage and probit models with all pooled tumors were comparable, however, BMD/BMDL ratios (above) and visual fit differ (see Figure 3a,b). Bold printed row indicates model chosen for extrapolation to humans.

TERA

24

ED_006319_00004789-00004